

WHAT IS CLAIMED IS:

- 5 1. A method of determining a title from a document image, comprising:
dividing the document image into minimal circumscribing rectangles which
contain a character image;
recognizing characters in said minimal circumscribing rectangles; and
determining a title of the document image based upon a likelihood of each of said
minimal circumscribing rectangles containing a title, said likelihood being determined
10 based upon information obtained during said character recognition.
2. The method of determining a title from a document image according to claim 1
wherein said likelihood is expressed in a sum of points based said information.
- 15 3. The method of determining a title from a document image according to claim 2
wherein said information includes characteristics on font.
4. The method of determining a title from a document image according to claim 2
wherein said font characteristics is determined on a frequency of a particular font type .
- 20 5. The method of determining a title from a document image according to claim 2
wherein said character recognition further includes an act of matching said characters with
a set of predetermined words, said predetermined words indicating said title.
- 25 6. The method of determining a title from a document image according to claim 5
wherein said information includes a result of said matching with said predetermined words.
7. The method of determining a title from a document image according to claim 2
wherein said information includes a number of said characters.
- 30

8. The method of determining a title from a document image according to claim 7 wherein said number of said characters is compared to a predetermined maximal threshold number.

5

Sub 23 p. The method of determining a title from a document image according to claim 2 wherein said information includes an assurance level of said character recognition.

10 9 10. The method of determining a title from a document image according to claim 9 wherein said assurance level is compared to a predetermined minimal threshold value.

Sub 24 11. The method of determining a title from a document image according to claim 2 wherein said information includes layout characteristics.

15 12. The method of determining a title from a document image according to claim 11 wherein said information includes centering, underlining and size.

20 13. The method of determining a title from a document image according to claim 2 wherein said information indicates whether or not said characters end in a noun form.

14. The method of determining a title from a document image according to claim 2 wherein said information indicates whether or not said characters end in a set of predetermined suffixes.

25 15. The method of determining a title from a document image according to claim 2 wherein said information includes a ratio between a length and a height of each of said circumscribing rectangles.

16. The method of determining a title from a document image according to claim 2 wherein said information includes a ratio between a summed width of said characters and a corresponding one of said circumscribing rectangles.

5 17. The method of determining a title from a document image according to claim 1 wherein said likelihood is adjusted according to a type of said image documents.

18. The method of determining a title from a document image according to claim 1 wherein said title is combined with a keyword.

10

19. A system for determining a title from a document image, comprising:
a character row area determination unit for dividing the document image into minimal circumscribing rectangles which contain a character image;
a character recognition unit connected to said character row area determination
15 unit for recognizing characters in said minimal circumscribing rectangles; and
a title evaluation point determination unit connected to said character recognition unit for determining a title of the document image based upon a likelihood of each of said minimal circumscribing rectangles containing a title, said likelihood being determined based upon information obtained during said character recognition.

20

20. The system for determining a title from a document image according to claim 19 wherein said title evaluation point determination unit determines said likelihood in terms of a sum of points based said information.

25

21. The system for determining a title from a document image according to claim 20 wherein said title evaluation point determination unit further comprises a font determination unit for generating information on font characteristics.

22. The system for determining a title from a document image according to claim 21 wherein said font determination unit determines said font characteristics based on a frequency of a particular font type.

5 23. The system for determining a title from a document image according to claim 20 wherein said title evaluation point determination unit further comprises a natural language analysis unit for matching said characters with a set of predetermined words, said predetermined words indicating said title.

10 24. The system for determining a title from a document image according to claim 23 wherein said natural language analysis unit generates a result of matching of said characters with said predetermined words.

15 ~~25. The system for determining a title from a document image according to claim 20 wherein said character recognition unit generates said information on a number of said characters.~~

20 26. The system for determining a title from a document image according to claim 25 wherein said title evaluation point determination unit compares said number of said characters to a predetermined maximal threshold number.

25 ~~27. The system for determining a title from a document image according to claim 20 wherein said character recognition unit generates said information on an assurance level of said character recognition.~~

28. The system for determining a title from a document image according to claim 27 wherein said title evaluation point determination unit compares said assurance level to a predetermined minimal threshold value.

29. The system for determining a title from a document image according to claim 20 wherein said title evaluation point determination unit further comprises a characteristics extraction unit for extracting layout characteristics.

5 ^{Sub 7} 30. The system for determining a title from a document image according to claim 29 wherein said extraction unit extracts said layout characteristics on centering, underlining and size.

10 31. The system for determining a title from a document image according to claim 23 wherein said natural language analysis unit generates said information indicating whether or not said characters end in a noun form.

15 32. The system for determining a title from a document image according to claim 23 wherein said natural language analysis unit generates said information indicating whether or not said characters end in a set of predetermined suffixes.

20 33. The system for determining a title from a document image according to claim 2 wherein said character row area determination unit generates said information on a ratio between a length and a height of each of said circumscribing rectangles.

25 34. The system for determining a title from a document image according to claim 2 wherein said character row area determination unit generates said information on a ratio between a summed width of said characters and a corresponding one of said circumscribing rectangles.

35. The system for determining a title from a document image according to claim 19 wherein said likelihood is adjusted according to a type of said image documents.

30 36. The system for determining a title from a document image according to claim 19 wherein said title is combined with a keyword.